



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> :  C07K 14/435, C12N 1/00, 1/19, 5/10, 15/12, 15/63		A1	(11) International Publication Number: <b>WO 00/34323</b>
			(43) International Publication Date: 15 June 2000 (15.06.00)
(21) International Application Number: PCT/US99/29404		(81) Designated States: JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(22) International Filing Date: 10 December 1999 (10.12.99)		Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(30) Priority Data: 09/210,330 11 December 1998 (11.12.98) US 09/444,338 19 November 1999 (19.11.99) US			
(71) Applicant: CLONTECH LABORATORIES, INC. [US/US]; 1020 East Meadow Drive, Palo Alto, CA 94303 (US).			
(72) Inventors: LUKYANOY, Sergey Anatolievich; ul. Golubinskaya, 13/1-161, Moscow (RU). FRADKOV, Arcady Fedorovich; ul. Dnepropetrovskaya, 35/2-14, Moscow, 113570 (RU). LABAS, Yulii Aleksandrovich; ul. Generala Tyuleneva, 35/416, Moscow, 117465 (RU). MATZ, Mikhail Vladimirovich; ul. Teplii stan, 7/2-28, Moscow, 117465 (RU).			
(74) Agent: ADLER, Benjamin, A.; McGregor & Adler, 8011 Candle Ln., Houston, TX 77071 (US).			
(54) Title: FLUORESCENT PROTEINS FROM NON-BIOLUMINESCENT SPECIES OF CLASS ANTHOZOA, GENES ENCODING SUCH PROTEINS AND USES THEREOF			
(57) Abstract  The present invention is directed to novel fluorescent proteins from non-bioluminescent organisms from the Class Anthozoa. Also disclosed are cDNAs encoding the fluorescent proteins.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		

**FLUORESCENT PROTEINS FROM NON-BIOLUMINESCENT SPECIES  
OF CLASS ANTHOZOA, GENES ENCODING SUCH PROTEINS AND  
USES THEREOF**

5

10

15 **Field of the Invention**

This invention relates to the field of molecular biology. More specifically, this invention relates to novel fluorescent proteins, cDNAs encoding the proteins and uses thereof.

20 **Description of the Related Art**

Fluorescence labeling is a particularly useful tool for marking a protein, cell, or organism of interest. Traditionally, a protein of interest is purified, then covalently conjugated to a fluorophore derivative. For *in vivo* studies, the protein-dye complex is 25 then inserted into cells of interest using micropipetting or a method of reversible permeabilization. The dye attachment and insertion steps, however, make the process laborious and difficult to control. An alternative method of labeling proteins of interest is to concatenate or 30 fuse the gene expressing the protein of interest to a gene expressing a marker, then express the fusion product. Typical markers for this

method of protein labeling include  $\beta$ -galactosidase, firefly luciferase and bacterial luciferase. These markers, however, require exogenous substrates or cofactors and are therefore of limited use for *in vivo* studies.

5 A marker that does not require an exogenous cofactor or substrate is the green fluorescent protein (GFP) of the jellyfish *Aequorea victoria*, a protein with an excitation maximum at 395 nm, a second excitation peak at 475 nm and an emission maximum at 510 nm. GFP is a 238-amino acid protein, with amino acids 65-67 involved  
10 in the formation of the chromophore.

Uses of GFP for the study of gene expression and protein localization are discussed in detail by Chalfie et al. in *Science* 263 (1994), 802-805, and Heim et al. in *Proc. Nat. Acad. Sci.* 91 (1994), 12501-12504. Additionally, Rizzuto et al. in *Curr. Biology* 5 (1995), 15 635-642, discuss the use of wild-type GFP as a tool for visualizing subcellular organelles in cells, while Kaether and Gerdes in *Febs Letters* 369 (1995), 267-271, report the visualization of protein transport along the secretory pathway using wild-type GFP. The expression of GFP in plant cells is discussed by Hu and Cheng in *Febs Letters* 369 (1995), 20 331-334, while GFP expression in *Drosophila* embryos is described by Davis et al. in *Dev. Biology* 170 (1995), 726-729.

Crystallographic structures of wild-type GFP and the mutant GFP S65T reveal that the GFP tertiary structure resembles a barrel (Ormö et al., *Science* 273 (1996), 1392-1395; Yang, et al., *Nature Biotechnol.* 14 (1996), 1246-1251). The barrel consists of beta sheets in a compact structure, where, in the center, an alpha helix containing the chromophore is shielded by the barrel. The compact structure makes GFP very stable under diverse and/or harsh conditions such as protease treatment, making GFP an extremely useful reporter in

general. However, the stability of GFP makes it sub-optimal for determining short-term or repetitive events.

A great deal of research is being performed to improve the properties of GFP and to produce GFP reagents useful and optimized for 5 a variety of research purposes. New versions of GFP have been developed, such as a "humanized" GFP DNA, the protein product of which has increased synthesis in mammalian cells (Haas, et al., *Current Biology* 6 (1996), 315-324; Yang, et al., *Nucleic Acids Research* 24 (1996), 4592-4593). One such humanized protein is "enhanced green 10 fluorescent protein" (EGFP). Other mutations to GFP have resulted in blue-, cyan- and yellow-green light emitting versions. Despite the great utility of GFP, however, other fluorescent proteins with properties similar to or different from GFP would be useful in the art. Novel 15 fluorescent proteins result in possible new colors, or produce pH-dependent fluorescence. Other benefits of novel fluorescent proteins include fluorescence resonance energy transfer (FRET) possibilities based on new spectra and better suitability for larger excitation.

The prior art is deficient in novel fluorescent proteins wherein the DNA coding sequences are known. The present invention 20 fulfills this long-standing need in the art.

## SUMMARY OF THE INVENTION

25 The present invention is directed to DNA sequences encoding fluorescent proteins selected from the group consisting of: (a) an isolated DNA from an organism from the Class Anthozoa which encodes a fluorescent protein; (b) an isolated DNA which hybridizes to the isolated DNA of (a) and which encodes a fluorescent protein; and

(c) an isolated DNA differing from the isolated DNAs of (a) and (b) in codon sequence due to the degeneracy of the genetic code and that encodes a fluorescent protein. Preferably, the DNA is isolated from a non-bioluminescent organism from Class Anthozoa. More preferably, 5 the DNA has the sequence shown in SEQ ID No. 55 and the fluorescent protein has the amino acid sequence shown in SEQ ID No. 56.

In another embodiment of the present invention, there is provided a vector capable of expressing the DNA of the present invention in a recombinant cell comprising said DNA and regulatory 10 elements necessary for expression of the DNA in the cell. Preferably, the DNA encodes a fluorescent protein having the amino acid sequence shown in SEQ ID No. 56.

In still another embodiment of the present invention, there is provided a host cell transfected with a vector of the present 15 invention, such that the host cell expresses a fluorescent protein. Preferably, the cell is selected from the group consisting of bacterial cells, mammalian cells, plant cells, insect cells and yeast cells. A representative example of bacterial cell is an *E. coli* cell.

The present invention is also directed to an isolated and 20 purified fluorescent protein coded for by DNA selected from the group consisting of: (a) isolated DNA from an organism from Class Anthozoa which encodes a fluorescent protein; (b) isolated DNA which hybridizes to the isolated DNA of (a) and which encodes a fluorescent protein; and (c) isolated DNA differing from the isolated DNAs of (a) and (b) in 25 codon sequence due to the degeneracy of the genetic code, and which encodes a fluorescent protein. Preferably, the protein has the amino acid sequence shown in SEQ ID No. 56.

The present invention is also directed to a DNA sequence encoding a fluorescent protein selected from the group consisting of:

(a) an isolated DNA which encodes a fluorescent protein, wherein said DNA is from an organism from Class Anthozoa and wherein said organism does not exhibit bioluminescence; (b) an isolated DNA which hybridizes to isolated DNA of (a) and which encodes a fluorescent protein; and (c) an isolated DNA differing from the isolated DNAs of (a) and (b) in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein. Preferably, the organism is from Sub-class Zoantharia, Order Corallimorpharia. More preferably, the organism is from Family Discosomatidae, Genus Discosoma. Even more preferably, the organism is *Discosoma* sp. "green". Most particularly, the present invention is drawn to a novel fluorescent protein from *Discosoma* sp. "green", dgFP512.

The present invention is further directed to an amino acid sequence which can be used as a basis for designing an oligonucleotide probe for identification of a DNA encoding a fluorescent protein by means of hybridizaton, wherein the amino acid sequence is selected from the group consisting of SEQ ID Nos. 3, 5, 8, 11, 12, 14. Preferably, such an oligonucleotide has a nucleotide sequence selected from the group consisting of SEQ ID Nos. 4, 6, 7, 9, 10, 13, 15, 16.

Other and further aspects, features, and advantages of the present invention will be apparent from the following description of the presently preferred embodiments of the invention given for the purpose of disclosure.

25

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the modified strategy of 3'-RACE used to isolate the target fragments. Sequences of the oligonucleotides used

are shown in Table 2. Dp1 and Dp2 are the degenerate primers used in the first and second PCR, respectively (see Tables 3 and 4 for the sequences of degenerate primers). In the case of *Discosoma* sp. "green", the first degenerate primer used was NGH (SEQ ID No. 4), and 5 the second degenerate primer used was NFP (SEQ ID No. 13) {Please confirm whether the degenerate primers used for dgFP512 are right. If not, please provide the right primers}

**Figure 2** shows the excitation and emission spectrum of the novel fluorescent protein from *Discosoma* sp. "green", dgFP512.

10

## DETAILED DESCRIPTION OF THE INVENTION

As used herein, the term "GFP" refers to the basic green 15 fluorescent protein from *Aequorea victoria*, including prior art versions of GFP engineered to provide greater fluorescence or fluoresce in different colors. The sequence of *Aequorea victoria* GFP (SEQ ID No. 54) has been disclosed in Prasher et al., *Gene* 111 (1992), 229-33.

As used herein, the term "EGFP" refers to mutant variant of 20 GFP having two amino acid substitutions: F64L and S65T (Heim et al., *Nature* 373 (1995), 663-664). The term "humanized" refers to changes made to the GFP nucleic acid sequence to optimize the codons for expression of the protein in human cells (Yang et al., *Nucleic Acids Research* 24 (1996), 4592-4593).

25 As used herein, the term "NFP" refers to novel fluorescent protein. Specifically, "NFP" refers to dgFP512 in the present invention.

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such

techniques are explained fully in the literature. See, e.g., Maniatis, Fritsch & Sambrook, "Molecular Cloning: A Laboratory Manual (1982); "DNA Cloning: A Practical Approach," Volumes I and II (D.N. Glover ed. 1985); "Oligonucleotide Synthesis" (M.J. Gait ed. 1984); "Nucleic Acid Hybridization" (B.D. Hames & S.J. Higgins eds. (1985)); "Transcription and Translation" (B.D. Hames & S.J. Higgins eds. (1984)); "Animal Cell Culture" (R.I. Freshney, ed. (1986)); "Immobilized Cells and Enzymes" (IRL Press, (1986)); B. Perbal, "A Practical Guide To Molecular Cloning" (1984).

10 A "vector" is a replicon, such as plasmid, phage or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment.

15 A "DNA molecule" refers to the polymeric form of deoxyribonucleotides (adenine, guanine, thymine, or cytosine) in either single stranded form or a double-stranded helix. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes.

20 A DNA "coding sequence" is a DNA sequence which is transcribed and translated into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' 25 (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic (e.g., mammalian) DNA, and synthetic DNA sequences. A polyadenylation signal and transcription termination sequence may be located 3' to the coding sequence.

As used herein, the term "hybridization" refers to the process of association of two nucleic acid strands to form an antiparallel duplex stabilized by means of hydrogen bonding between residues of the opposite nucleic acid strands.

5 The term "oligonucleotide" refers to a short (under 100 bases in length) nucleic acid molecule.

"DNA regulatory sequences", as used herein, are transcriptional and translational control sequences, such as promoters, enhancers, polyadenylation signals, terminators, and the like, that 10 provide for and/or regulate expression of a coding sequence in a host cell.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining 15 the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a 20 transcription initiation site, as well as protein binding domains responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes. Various promoters, including inducible promoters, may be used to drive the various vectors of the present invention.

25 As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence.

A cell has been "transformed" or "transfected" by exogenous or heterologous DNA when such DNA has been introduced

inside the cell. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to 5 eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter 10 cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

A "heterologous" region of the DNA construct is an 15 identifiable segment of DNA within a larger DNA molecule that is not found in association with the larger molecule in nature. Thus, when the heterologous region encodes a mammalian gene, the gene will usually be flanked by DNA that does not flank the mammalian genomic DNA in the genome of the source organism. In another example, 20 heterologous DNA includes coding sequence in a construct where portions of genes from two different sources have been brought together so as to produce a fusion protein product. Allelic variations or naturally-occurring mutational events do not give rise to a heterologous region of DNA as defined herein.

25 As used herein, the term "reporter gene" refers to a coding sequence attached to heterologous promoter or enhancer elements and whose product may be assayed easily and quantifiably when the construct is introduced into tissues or cells.

The amino acids described herein are preferred to be in the "L" isomeric form. The amino acid sequences are given in one-letter code (A: alanine; C: cysteine; D: aspartic acid; E: glutamic acid; F: phenylalanine; G: glycine; H: histidine; I: isoleucine; K: lysine; L: leucine; 5 M: methionine; N: asparagine; P: proline; Q: glutamine; R: arginine; S: serine; T: threonine; V: valine; W: tryptophane; Y: tyrosine; X: any residue). NH<sub>2</sub> refers to the free amino group present at the amino terminus of a polypeptide. COOH refers to the free carboxy group present at the carboxy terminus of a polypeptide. In keeping with 10 standard polypeptide nomenclature, *J Biol. Chem.*, 243 (1969), 3552-59 is used.

The present invention is directed to an isolated DNA selected from the group consisting of: (a) isolated DNA from an organism from the Class Anthozoa which encodes a fluorescent 15 protein; (b) isolated DNA which hybridizes to isolated DNA of (a) and which encodes a fluorescent protein; and (c) isolated DNA differing from the isolated DNAs of (a) and (b) in codon sequence due to the degeneracy of the genetic code, and which encodes a fluorescent protein. Preferably, the DNA has the sequence shown in SEQ ID No. 55 20 and the fluorescent protein has the amino acid sequence shown in SEQ ID No. 56.

In another embodiment of the present invention, there is provided a vector capable of expressing the DNA of the present invention in a recombinant cell comprising said DNA and regulatory 25 elements necessary for expression of the DNA in the cell. Specifically, the DNA encodes a fluorescent protein having the amino acid sequence shown in SEQ ID No. 56.

In still another embodiment of the present invention, there is provided a host cell transfected with the vector of the present

invention, which expresses a fluorescent protein of the present invention. Preferably, the cell is selected from the group consisting of bacterial cells, mammalian cells, plant cells, insect cells and yeast cells. A representative example of bacterial cell is an *E. coli* cell.

5 The present invention is also directed to a DNA sequence encoding a fluorescent protein selected from the group consisting of: (a) an isolated DNA which encodes a fluorescent protein, wherein said DNA is from an organism from Class Anthozoa and wherein said organism does not exhibit bioluminescence; (b) an isolated DNA which 10 hybridizes to isolated DNA of (a) and which encodes a fluorescent protein; and (c) an isolated DNA differing from the isolated DNAs of (a) and (b) in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein. Preferably, the organism is from Sub-class Zoantharia, Order Corallimorpharia. More preferably, 15 the organism is from Family Discosomatidae, Genus Discosoma. Most preferably, the organism is *Discosoma* sp. "green".

The present invention is also directed to an isolated and purified fluorescent protein coded for by DNA selected from the group consisting of: (a) an isolated protein encoded by a DNA which encodes 20 a fluorescent protein wherein said DNA is from an organism from Class Anthozoa and wherein said organism does not exhibit bioluminescence; (b) an isolated protein encoded by a DNA which hybridizes to isolated DNA of (a); and (c) an isolated protein encoded by a DNA differing from the isolated DNAs of (a) and (b) in codon sequence due to 25 degeneracy of the genetic code. Preferably, the isolated and purified fluorescent protein is dgFP512.

The present invention is further directed to an amino acid sequence which can be used as a basis for designing an oligonucleotide probe for identification of a DNA encoding a fluorescent protein by

means of hybridizaton, wherein the amino acid sequence is selected from the group consisting of SEQ ID Nos. 3, 5, 8, 11, 12, 14. Preferably, such an oligonucleotide has a nucleotide sequence selected from the group consisting of SEQ ID Nos. 4, 6, 7, 9, 10, 13, 15, 16 and is used as 5 a primer in polymerase chain reaction. Alternatively, it can be used as a probe for hybridization screening of the cloned genomic or cDNA library.

10 The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion.

### **EXAMPLE 1**

#### **Biological Material**

15 Novel fluorescent proteins were identified from several genera of Anthozoa which do not exhibit any bioluminescence but have fluorescent color as observed under usual white light or ultraviolet light. Six species were chosen (see Table 1).

20

TABLE 1

Anthozoa Species Used in This Study

Species	Area of Origination	Fluorescent Color
<i>Anemonia majano</i>	Western Pacific	bright green tentacle tips
<i>Clavularia</i> sp.	Western Pacific	bright green tentacles and oral disk
<i>Zoanthus</i> sp.	Western Pacific	green-yellow tentacles and oral disk
<i>Discosoma</i> sp. "red"	Western Pacific	orange-red spots oral disk
<i>Discosoma</i> sp. "striata"	Western Pacific	blue-green stripes on oral disk
<i>Discosoma</i> sp. "magenta"	Western Pacific	faintly purple oral disk
<i>Discosoma</i> sp. "green"	Western Pacific	green spots on oral disk
<i>Anemonia sulcata</i>	Mediterranean	purple tentacle tips

**EXAMPLE 2****cDNA Preparation**

Total RNA was isolated from the species of interest according to the protocol of Chomczynski and Sacchi (Chomczynski P., et al., *Anal. Biochem.* 162 (1987), 156-159). First-strand cDNA was synthesized starting with 1-3 µg of total RNA using SMART PCR cDNA synthesis kit (CLONTECH) according to the provided protocol with the only alteration being that the "cDNA synthesis primer" provided in the kit was replaced by the primer TN3 (5'- CGCAGTCGACCG(T)<sub>13</sub>, SEQ ID No. 1) (Table 2). Amplified cDNA samples were then prepared as described in the protocol provided except the two primers used for PCR were the TS primer (5'-AAGCAGTGGTATCAACGCAGAGT, SEQ ID No. 2) (Table 2) and the TN3 primer (Table 2), both in 0.1 µM concentration. Twenty to twenty-five PCR cycles were performed to amplify a cDNA sample. The amplified cDNA was diluted 20-fold in water and 1 µl of this dilution was used in subsequent procedures.

**TABLE 2****Oligos Used in cDNA Synthesis and RACE**

5 TN3: 5'-CGCAGTCGACCG(T)<sub>13</sub>  
(SEQ ID No. 1)

10 T7-TN3: 5'-GTAATACGACTCACTATAGGGCCGCAGTCGACCG(T)<sub>13</sub>  
(SEQ ID No. 17)

15 TS-primer: 5'-AAGCAGTGGTATCAACGCAGAGT  
(SEQ ID No. 2)

20 T7-TS:  
15 5'-GTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT  
(SEQ ID No. 18)

25 T7: 5'-GTAATACGACTCACTATAGGGC  
(SEQ ID No. 19)

TS-oligo 5'-AAGCAGTGGTATCAACGCAGAGTACGCrGrGrG  
(SEQ ID No. 53)

**EXAMPLE 3**Oligo Design

To isolate fragments of novel fluorescent protein cDNAs, PCR using degenerate primers was performed. Degenerate primers were designed to match the sequence of the mRNAs in regions that were predicted to be the most invariant in the family of fluorescent proteins. Four such stretches were chosen (Table 3) and variants of degenerate primers were designed. All such primers were directed to the 3'-end of mRNA. All oligos were gel-purified before use. Table 2 shows the oligos used in cDNA synthesis and RACE.

TABLE 3

5 Key Amino Acid Stretches and Corresponding Degenerate Primers Used for Isolation of Fluorescent Proteins

Stretch Position according to A. victoria GFP (7)	Amino Acid Sequence of the Key Stretch	Degenerated Primer Name and Sequence
20-25	GXVN <sup>GH</sup> (SEQ ID No. 3)	NGH: 5'- GA(C,T) GGC TGC GT(A,T,G,C) AA(T,C) GG(A,T,G) CA (SEQ ID No. 4)
31-35	GEGEG (SEQ ID No. 5)	GEGa: 5'- GTT ACA GGT GA(A,G) GG(A,C) GA(A,G) GG (SEQ ID No. 6) GEGb: 5'- GTT ACA GGT GA(A,G) GG(T,G) GA(A,G) GG (SEQ ID No. 7) GN <sup>GA</sup> : 5'- GTT ACA GGT GA(A,G) GG(A,C) AA(C,T) GG (SEQ ID No. 9) GN <sup>GB</sup> : 5'- GTT ACA GGT GA(A,G) GG(T,G) AA(C,T) GG (SEQ ID No. 10)
127-131	GMN <sup>FP</sup> (SEQ ID No. 11) GVN <sup>FP</sup> (SEQ ID No. 12)	NFP: 5' TTC CA(C,T) GGT (G,A)TG AA(C,T) TT(C,T) CC (SEQ ID NO. 13)
134-137	GPV <sup>M</sup> (SEQ ID No. 14)	PVMa: 5' CCT GCC (G,A)A(C,T) GGT CC(A,T,G,C) GT(A,C) ATG (SEQ ID NO. 15) PVMb: 5' CCT GCC (G,A)A(C,T) GGT CC(A,T,G,C) GT(G,T) ATG (SEQ ID NO. 16)

**EXAMPLE 4**Isolation of 3'-cDNA Fragments of nFPs

The modified strategy of 3'-RACE was used to isolate the target fragments (see Figure 1). The RACE strategy involved two consecutive PCR steps. The first PCR step involved a first degenerate primer (Table 4) and the T7-TN3 primer (SEQ ID No. 17) which has a 3' portion identical to the TN3 primer used for cDNA synthesis (for sequence of T7-TN3, Table 2). The reason for substituting the longer T7-TN3 primer in this PCR step was that background amplification which occurred when using the shorter TN3 primer was suppressed effectively, particularly when the T7-TN3 primer was used at a low concentration (0.1  $\mu$ M) (Frohman et al., (1998) *PNAS USA*, 85, 8998-9002). The second PCR step involved the TN3 primer (SEQ ID No. 1, Table 2) and a second degenerate primer (Table 4).

TABLE 4

Combinations of Degenerate Primers for First and Second PCR Resulting in Specific Amplification of 3'-Fragments of nFP cDNA

Species	First Degenerate Primer	Second Degenerate Primer
<i>Anemonia majano</i>	NGH (SEQ ID No. 4)	GNGb (SEQ ID No. 10)
<i>Clavularia</i> sp.	NGH (SEQ ID No. 4)	GEGa (SEQ ID No. 6)
<i>Zoanthus</i> sp.	NGH (SEQ ID No. 4)	GEGa (SEQ ID No. 6)
<i>Discosoma</i> sp. "red"	NGH (SEQ ID No. 4)	GEGa (SEQ ID No. 6), NFP (SEQ ID No. 13) or PVMb (SEQ ID No. 16)
<i>Discosoma striata</i>	NGH (SEQ ID No. 4)	NFP (SEQ ID No. 13)
<i>Anemonia sulcata</i>	NGH (SEQ ID No. 4)	GEGa (SEQ ID No. 6) or NFP (SEQ ID No. 13)

5

The first PCR reaction was performed as follows: 1  $\mu$ l of 20-fold  
 10 dilution of the amplified cDNA sample was added into the reaction  
 mixture containing 1X Advantage KlenTaq Polymerase Mix with  
 provided buffer (CLONTECH), 200  $\mu$ M dNTPs, 0.3  $\mu$ M of first degenerate

primer (Table 4) and 0.1  $\mu$ M of T7-TN3 (SEQ ID No. 17) primer in a total volume of 20  $\mu$ l. The cycling profile was (Hybaid OmniGene Thermocycler, tube control mode): 1 cycle for 95°C, 10 sec.; 55°C, 1 min.; 72°C, 40 sec; 24 cycles for 95°C, 10 sec.; 62°C, 30 sec.; 72°C, 40 sec. The reaction was then diluted 20-fold in water and 1  $\mu$ l of this dilution was added to a second PCR reaction, which contained 1X Advantage KlenTaq Polymerase Mix with the buffer provided by the manufacturer (CLONTECH), 200  $\mu$ M dNTPs, 0.3  $\mu$ M of the second degenerate primer (Table 4) and 0.1  $\mu$ M of TN3 primer. The cycling profile was (Hybaid OmniGene Thermocycler, tube control mode): 1 cycle for 95°C, 10 sec.; 55°C (for GEG/GNG or PVM) or 52°C (for NFP), 1 min.; 72°C, 40 sec; 13 cycles for 95°C, 10 sec.; 62°C (for GEG/GNG or PVM) or 58°C (for NFP), 30 sec.; 72°C, 40 sec. The product of PCR was cloned into PCR-Script vector (Stratagene) according to the manufacturer's protocol.

Different combinations of degenerate primers were tried in the first and second PCR reactions on the DNA from each species until a combination of primers was found that resulted in specific amplification--meaning that a pronounced band of expected size (about 650-800 bp for NGH and GEG/GNG and 350-500 bp for NFP and PVM--sometimes accompanied by a few minor bands) was detected on agarose gel after two PCR reactions. The primer combinations of choice for different species of the Class Anthozoa are listed in Table 4. Some other primer combinations also resulted in amplification of fragments of correct size, but the sequence of these fragments showed no homology to the other fluorescent proteins identified or to *Aequorea victoria* GFP.

**EXAMPLE 5****Obtaining Full-Length cDNA Copies**

Upon sequencing the obtained 3'-fragments of novel fluorescent protein cDNAs, two nested 5'-directed primers were synthesized for cDNA (Table 5), and the 5' ends of the cDNAs were then amplified using two consecutive PCRs. In the next PCR reaction, the novel approach of "step-out PCR" was used to suppress background amplification. The step-out reaction mixture contained 1x Advantage 5 KlenTaq Polymerase Mix using buffer provided by the manufacturer (CLONTECH), 200  $\mu$ M dNTPs, 0.2  $\mu$ M of the first gene-specific primer (see Table 5), 0.02  $\mu$ M of the T7-TS primer (SEQ ID No. 18), 0.1  $\mu$ M of T7 primer (SEQ ID No. 19) and 1  $\mu$ l of the 20-fold dilution of the amplified cDNA sample in a total volume of 20  $\mu$ l. The cycling profile 10 was (Hybaid OmniGene Thermocycler, tube control mode): 23-27 cycles for 95°C, 10 sec.; 60°C, 30 sec.; 72°C, 40 sec. The product of amplification was diluted 50-fold in water and one  $\mu$ l of this dilution was added to the second (nested) PCR. The reaction contained 1X 15 Advantage KlenTaq Polymerase Mix with provided buffer (CLONTECH), 200  $\mu$ M dNTPs, 0.2  $\mu$ M of the second gene-specific primer and 0.1  $\mu$ M of TS primer (SEQ ID No. 2) in a total volume of 20  $\mu$ l. The cycling profile was (Hybaid OmniGene Thermocycler, tube control mode): 12 cycles for 95°C, 10 sec.; 60°C, 30 sec.; 72°C, 40 sec. The product of amplification was then cloned into pAtlas vector (CLONTECH) according 25 to the manufacturer's protocol.

TABLE 5

## Gene-Specific Primers Used for 5'-RACE

Species	First Primer	Second (Nested) Primer
Anemonia majano	5'-GAAATAGTCAGGCATACTGGT (SEQ ID No. 20)	5'-GTCAGGCATAC TGGTAGGAT (SEQ ID No. 21)
Clavularia sp.	5'-CTTGAAATAGTCTGCTATATC (SEQ ID No. 22)	5'-TCTGCTATATC GTCTGGGT (SEQ ID No. 23)
Zoanthus sp.	5'- GTTCTTGAAATAGTCTACTATGT (SEQ ID No. 24)	5'-GTCTACTATGTCTT GAGGAT (SEQ ID No. 25)
Discosoma sp. "red"	5'-CAAGCAAATGGCAAAGGTC (SEQ ID No. 26)	5'-CGGTATTGTGGCC TTCGTA (SEQ ID No. 27)
Discosoma striata	5'-TTGTCTTCTTCTGCACAAC (SEQ ID No. 28)	5'-CTGCACAAACGG GTCCAT (SEQ ID No. 29)
Anemonia sulcata	5'-CCTCTATCTTCATTTCTGC (SEQ ID No. 30)	5'-TATCTTCATTTCT GCGTAC (SEQ ID No. 31)
Discosoma sp. "magenta"	5'-TTCAGCACCCATCACGAG (SEQ ID No. 32)	5'-ACGCTCAGAGCTG GGTTCC (SEQ ID No. 33)
Discosoma sp. "green"	5'-CCCTCAGCAATCCATCACGTT (SEQ ID No. 34)	5'-ATTATCTCAGTGGAA TGGTTC (SEQ ID No. 35)

**EXAMPLE 6**Expression of nFP in *E.coli*

5           To prepare a DNA construct for novel fluorescent protein expression, two primers were synthesized for each cDNA: a 5'-directed "downstream" primer with the annealing site located in the 3'-UTR of the cDNA and a 3'-directed "upstream" primer corresponding to the site of translation start site (not including the first ATG codon) (Table 10 6). Primers with SEQ ID Nos. 51 and 52 were the primers used to prepare the dgFP512 DNA. Both primers had 5'-heels coding for a site for a restriction endonuclease; in addition, the upstream primer was designed so as to allow the cloning of the PCR product into the pQE30 vector (Qiagen) in such a way that resulted in the fusion of reading 15 frames of the vector-encoded 6xHis-tag and nFP. The PCR was performed as follows: 1  $\mu$ l of the 20-fold dilution of the amplified cDNA sample was added to a mixture containing 1x Advantage KlenTaq Polymerase Mix with buffer provided by the manufacturer (CLONTECH), 200  $\mu$ M dNTPs, 0.2  $\mu$ M of upstream primer and 0.2  $\mu$ M of downstream 20 primer, in a final total volume of 20  $\mu$ l. The cycling profile was (Hybaid OmniGene Thermocycler, tube control mode): 23-27 cycles for 95°C, 10 sec.; 60°C, 30 sec.; 72°C, 40 sec. The product of this amplification step was purified by phenol-chlorophorm extraction and ethanol precipitation and then cloned into pQE30 vector using restriction 25 endonucleases corresponding to the primers' sequence according to standard protocols.

All plasmids were amplified in XL-1 blue *E. coli* and purified by plasmid DNA miniprep kits (CLONTECH). The recombinant clones were selected by colony color, and grown in 3 ml of LB medium

(supplemented with 100  $\mu$ g/ml of ampicillin) at 37°C overnight. 100  $\mu$ l of the overnight culture was transferred into 200 ml of fresh LB medium containing 100  $\mu$ g/ml of ampicillin and grown at 37°C, 200 rpm up to OD<sub>600</sub> 0.6-0.7. 1 mM IPTG was then added to the culture and 5 incubation was allowed to proceed at 37°C for another 16 hours. The cells were harvested and recombinant protein, which incorporated 6x His tags on the N-terminus, was purified using TALON™ metal-affinity resin according to the manufacturer's protocol (CLONTECH).

TABLE 6

Primers Used to Obtain Full Coding Region of nFPs for Cloning into Expression Construct

Species	Upstream Primer	Downstream Primer
Anemonia majano	5' - <u>acatggatccgctttcaaaca</u> agtttatc (SEQ ID No. 36) BamHI	5'-tagtact <u>c gagcttattcgta</u> tttcagtcaa (SEQ ID No. 37) Xhol
Clavularia sp.	L: 5'- <u>acatggatccaaacatttttga</u> gaaacg (SEQ ID No. 38) BamHI S: 5'- <u>acatggatccaaagctctaacc</u> accatg (SEQ ID No. 39) BamHI	5'-tagtact <u>c gagcaacacaa</u> accctcagacaa (SEQ ID No. 40) Xhol
Zoanthus sp.	5'- <u>acatggatccgctcagtcaaag</u> cacgg (SEQ ID No. 41) BamHI	5'-tagtact <u>c gaggttggaaactacat</u> tcttatca (SEQ ID No. 42) Xhol
Discosoma sp. "red"	5'- <u>acatggatccaggcgtttccaaagaat</u> gttatac (SEQ ID No. 43) BamHI	5'-tagtact <u>c gagggccaagttc</u> agccta (SEQ ID No. 44) Xhol
Discosoma striata	5'- <u>acatggatccaggcgttccaaagagtgt</u> (SEQ ID No. 45) BamHI	5'-tag <u>c gagcctatcatgcctc</u> gtcacct (SEQ ID No. 46) SacI
Anemonia sulcata	5'- <u>acatggatccgcttcctttaaagaagact</u> (SEQ ID No. 47) BamHI	5'-tagtact <u>c gagtcctgggagc</u> ggcttg (SEQ ID No. 48) Xhol
Discosoma sp. "magenta"	5'- <u>acatggatccaggcgttccaaagaatgtat</u> (SEQ ID No. 49) BamHI	5'-tagtact <u>c gagggcattacg</u> ctaatac (SEQ ID No. 50) Xhol
Discosoma sp. "green"	5'- <u>acatggatccaggcgtcactttaaagaagaaat</u> (SEQ ID No. 51)	5'-tagtact <u>c gagattcggttaat</u> gccttg (SEQ ID No. 52)

**EXAMPLE 7****Novel Fluorescent Proteins and cDNAs Encoding the Proteins**

One of the full-length cDNAs encoding fluorescent proteins found is described herein (dgFP512). The nucleic acid sequence and deduced amino acid sequence are SEQ ID Nos. 55 and 56, respectively. The spectral properties of dgFP512 is listed in Table 7, and the emission and excitation spectra for the dgFP512 is shown in Figure 2.

10

**TABLE 7****Spectral Properties of the Isolated dgFP512**

Species:	Discosoma sp. "green"	Max. Extinction Coefficient:	20,360
15 nFP Name:	dgFP512	Quantum Yield	0.24
Absorbance Max. (nm):	502	Relative Brightness:*	0.21
20 Emission Max. (nm):	512		

\*relative brightness is extinction coefficient multiplied by quantum yield divided by the same value for *A. victoria* GFP.

25

Any patents or publications mentioned in this specification are indicative of the levels of those skilled in the art to which the invention pertains. These patents and publications are incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference

One skilled in the art will appreciate readily that the present invention is adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those objects and ends inherent therein. The present examples, along with the methods, procedures, 5 treatments, molecules, and specific compounds described herein, are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes to the methods and compounds, and other uses, will occur to those skilled in the art and are encompassed within the spirit of the invention 10 as defined by the scope of the claims.

**WHAT IS CLAIMED IS:**

1. A DNA sequence encoding a fluorescent protein selected from the group consisting of:

5 (a) an isolated DNA which encodes a fluorescent protein, wherein said DNA is from an organism from a Class Anthozoa and wherein said organism does not exhibit bioluminescence;

(b) an isolated DNA which hybridizes to isolated DNA of (a) above and which encodes a fluorescent protein; and

10 (c) an isolated DNA differing from the isolated DNAs of (a) and (b) above in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein.

2. The DNA sequence of claim 1, wherein said organism 15 is from Sub-class Zoantharia.

3. The DNA sequence of claim 2, wherein said organism is from Order Corallimorpharia.

20 4. The DNA sequence of claim 3, wherein said organism is from Family Discosomatidae.

5. The DNA sequence of claim 4, wherein said organism is from Genus Discosoma.

25 6. The DNA sequence of claim 5, wherein said organism is *Discosoma* sp. "green".

7. A DNA sequence encoding a fluorescent protein selected from the group consisting of:

(a) an isolated DNA which encodes a fluorescent protein having a nucleotide sequence shown in SEQ ID No. 55;

5 (b) an isolated DNA which hybridizes to isolated DNA of (a) above and which encodes a fluorescent protein; and

(c) an isolated DNA differing from the isolated DNAs of (a) and (b) above in codon sequence due to degeneracy of the genetic code, and which encodes a fluorescent protein.

10

8. — The DNA of claim 7, wherein said DNA encodes a fluorescent protein having an amino acid sequence shown in SEQ ID No. 56.

15

9. A vector capable of expressing the DNA of claim 1 in a recombinant cell, said vector comprising said DNA of claim 1 and regulatory elements necessary for expression of the DNA in the cell.

20

10. The vector of claim 9, wherein said DNA encodes a fluorescent protein having the amino acid sequence shown in SEQ ID No. 56.

11. A host cell transfected with the vector of claim 9, wherein said cell is capable of expressing a fluorescent protein.

25

12. The host cell of claim 11, wherein said cell is selected from the group consisting of bacterial cells, mammalian cells, plant cell, yeast and insect cells.

13. The host cell of claim 12, wherein said bacterial cell is an *E. coli* cell.

14. An isolated and purified fluorescent protein coded for 5 by DNA selected from the group consisting of:

(a) an isolated DNA which encodes a fluorescent protein from an organism from Class Anthozoa, wherein said organism does not exhibit bioluminescence;

10 (b) an isolated DNA which hybridizes to isolated DNA of (a) above and which encodes a fluorescent protein; and

(c) an isolated DNA differing from the isolated DNAs of (a) and (b) above in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein.

15 15. The isolated and purified fluorescent protein of claim 14, wherein said organism is from Sub-class Zoantharia.

16. The isolated and purified fluorescent protein of claim 15, wherein said organism is from Order Corallimorpharia.

20 17. The isolated and purified fluorescent protein of claim 16, wherein said organism is from Family Discosomatidae.

18. The isolated and purified fluorescent protein of claim 25 17, wherein said organism is from Genus Discosoma.

19. The isolated and purified fluorescent protein of claim 18, wherein said organism is *Discosoma* sp. "green".

20. An isolated and purified fluorescent protein coded for by DNA selected from the group consisting of:

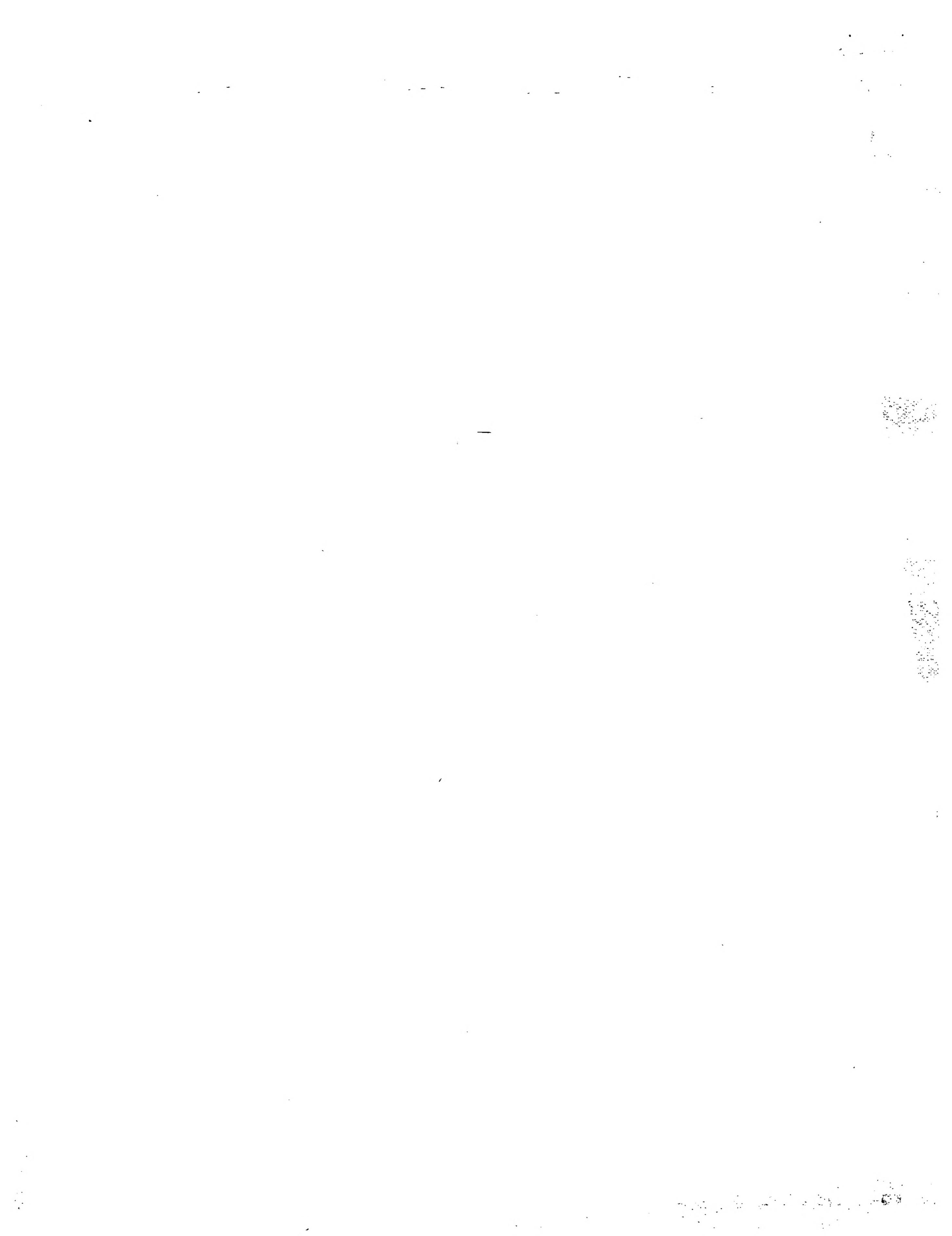
- (a) isolated DNA which encodes a fluorescent protein having an amino acid sequence shown in SEQ ID No. 56;
- 5 (b) isolated DNA which hybridizes to isolated DNA of (a) above and which encodes a fluorescent protein; and
- (c) isolated DNA differing from said isolated DNAs of (a) and (b) above in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein.

10

21. The isolated and purified fluorescent protein of claim 20, wherein said protein is dgFP512.

22. An amino acid sequence which can be used as a basis 15 for designing an oligonucleotide probe for identification of a DNA encoding a fluorescent protein by means of hybridizaton, wherein said sequence is selected from the group consisting of SEQ ID Nos. 3, 5, 8, 11, 12, 14.

20 23. The amino acid sequence of claim 22, wherein said oligonucleotide has a nucleotide sequence selected from the group consisting of SEQ ID Nos. 4, 6, 7, 9, 10, 13, 15, 16



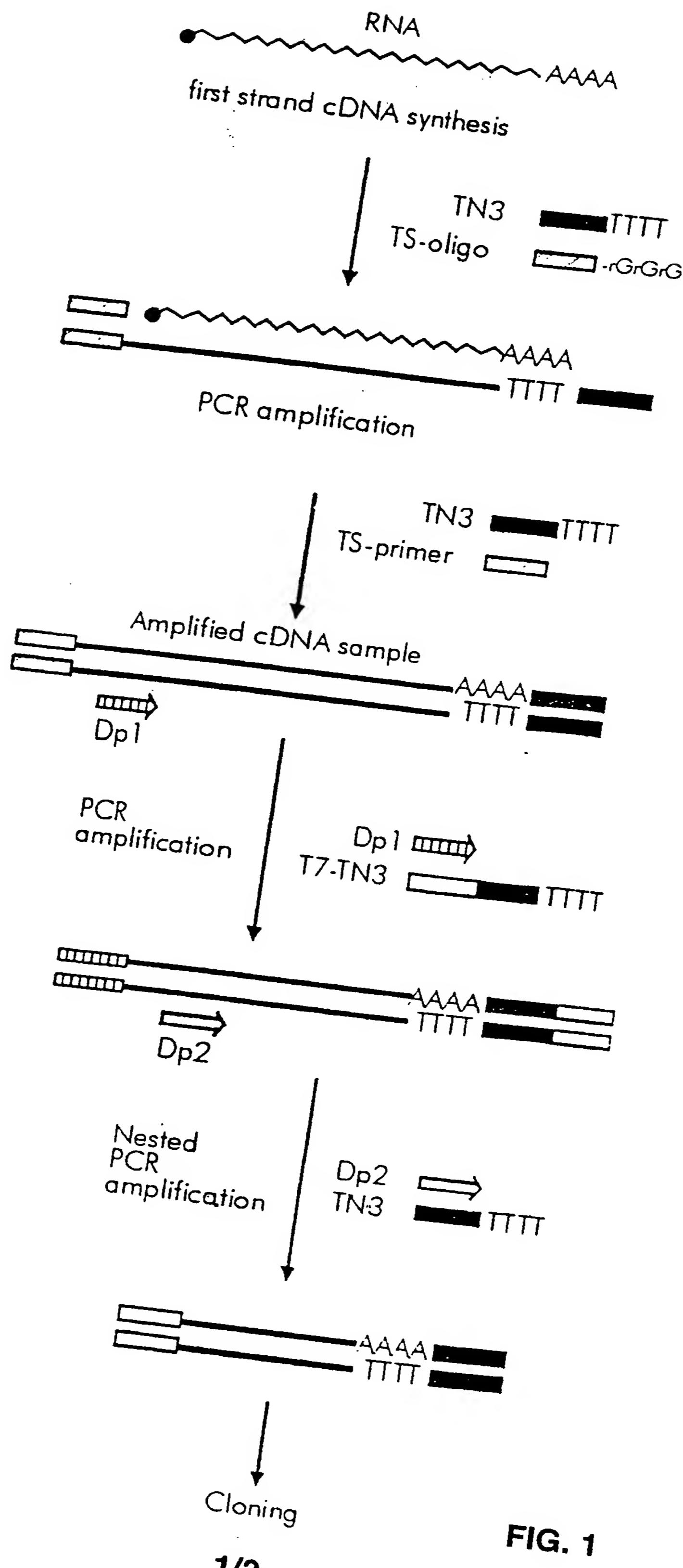


FIG. 1



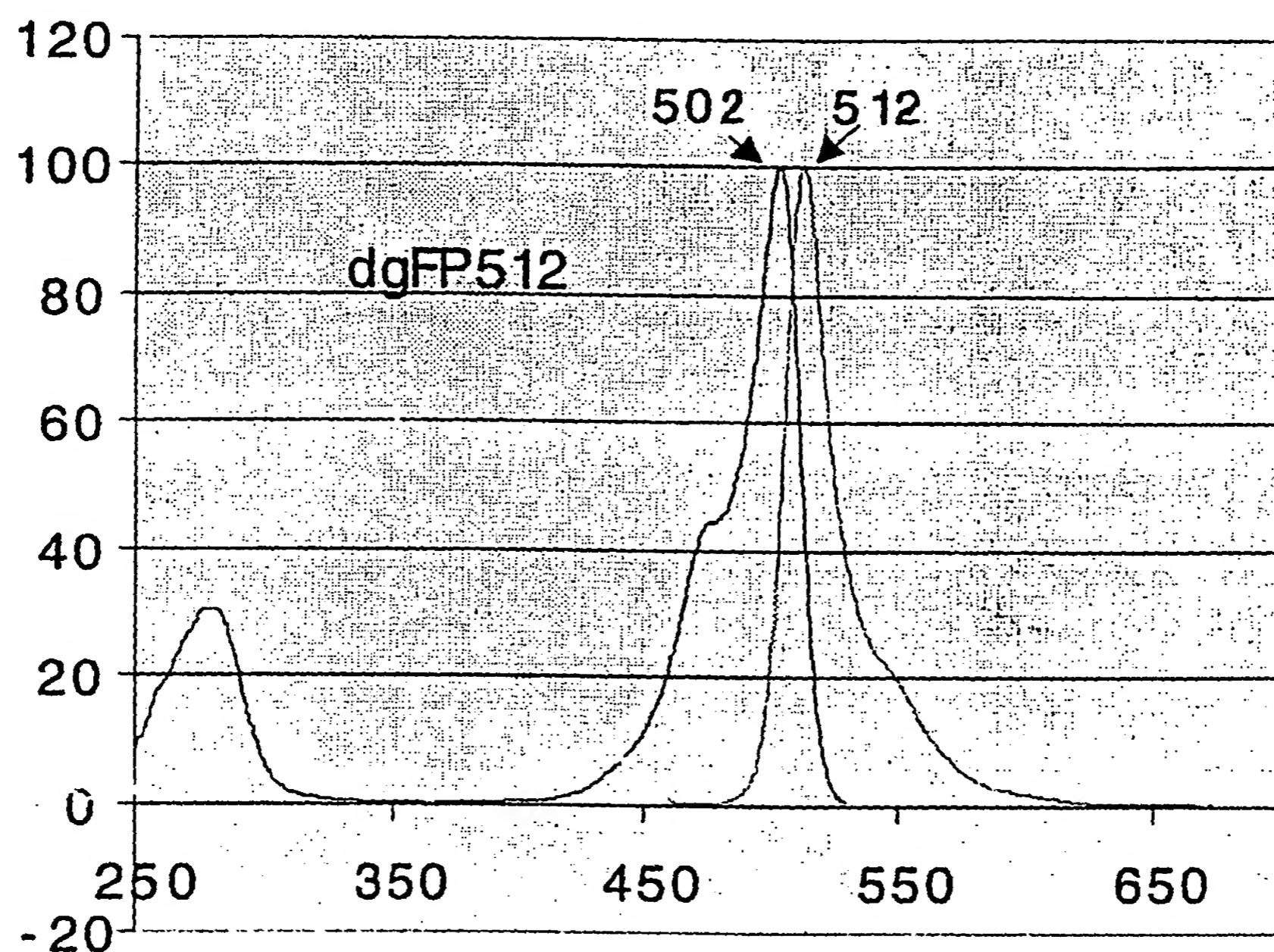
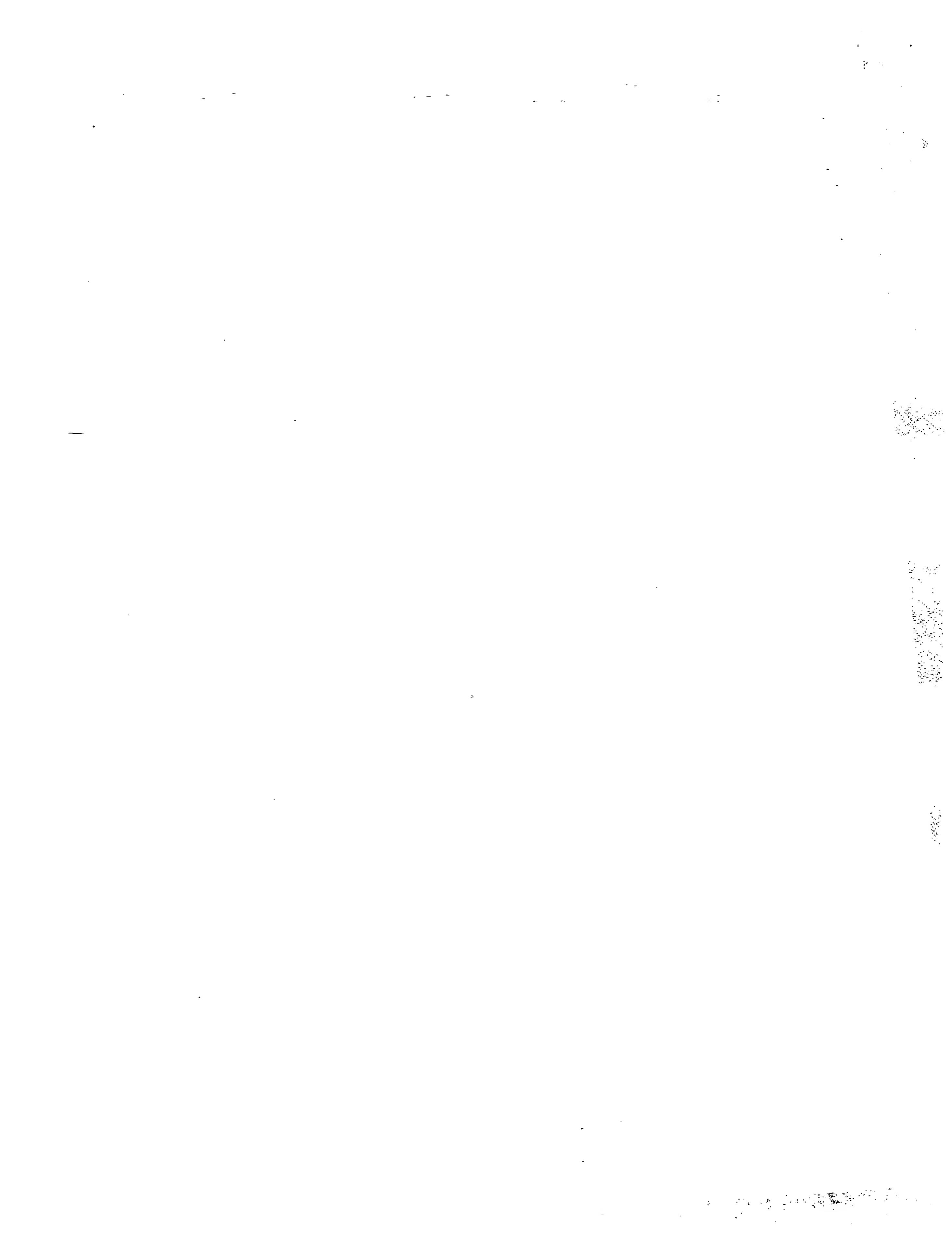


FIG. 2



## SEQUENCE LISTING

<110> Lukyanov, Sergey A.  
Labas, Yulii A.  
Matz, Mikhail V.  
Fradkov, Arcady F.

<120> Fluorescent Proteins from Non-Bioluminescent  
Species of Class Anthozoa, Genes Encoding Such  
Proteins and Uses Thereof

<130> D6196D8PCT

<140> 09/444,338

<141> 1999-11-19

<150> 09/210,330

<151> 1998-12-11

<160> 56

<210> 1

<211> 25

<212> DNA

<213> artificial sequence

<220>

<221> primer\_bind

<223> primer TN3 used in cDNA synthesis and RACE

<400> 1

cgcagtcgac cgttttttt ttttt

25

<210> 2

<211> 23

<212> DNA

<213> artificial sequence

<220>

<221> primer\_bind

<223> primer TS used in cDNA synthesis and RACE

<400> 2

aagcagtggat atcaacgcag agt

23

<210> 3



<211> 6  
<212> PRT  
<213> *Aequorea victoria*  
<220>  
<222> 21  
<223> amino acid sequence of a key stretch on which  
primer NGH is based; Xaa at position 21  
represents  
unknown  
<400> 3

Gly Xaa Val Asn Gly His

5

<210> 4  
<211> 20  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<222> 12  
<223> primer NGH used for isolation of fluorescent  
protein; n at position 12 represents any of the  
four bases  
<400> 4

gayggctgcg tnaayggdca

20

<210> 5  
<211> 5  
<212> PRT  
<213> *Aequorea victoria*  
<220>  
<222> 31...35  
<223> amino acid sequence of a key stretch on which  
primers GEGa and GEGb are based  
<400> 5

Gly Glu Gly Glu Gly

5



<210> 6  
<211> 20  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> primer GEGA used for isolation of fluorescent protein  
<400> 6

gttacaggtg arggmgargg 20

<210> 7  
<211> 20  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> primer GEGb used for isolation of fluorescent protein  
<400> 7

gttacaggtg arggkgargg 20

<210> 8  
<211> 5  
<212> PRT  
<213> *Aequorea victoria*  
<220>  
<222> 31...35  
<223> amino acid sequence of a key stretch on which primers GNGa and GNGb are based  
<400> 8

Gly Glu Gly Asn Gly

5

<210> 9  
<211> 20



<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> primer GNGa used for isolation of fluorescent protein  
<400> 9

gttacaggtg arggmaaygg 20

<210> 10  
<211> 20  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> primer GNGb used for isolation of fluorescent protein  
<400> 10

gttacaggtg arggkaaygg 20

<210> 11  
<211> 5  
<212> PRT  
<213> *Aequorea victoria*  
<220>  
<222> 127..131  
<223> amino acid sequence of a key stretch on which primer NFP is based  
<400> 11

Gly Met Asn Phe Pro

5

<210> 12  
<211> 5  
<212> PRT  
<213> *Aequorea victoria*  
<220>



<222> 127...131  
 <223> amino acid sequence of a key stretch on which  
 <400> primer NFP is based  
 12

Gly Val Asn Phe Pro

5

<210> 13  
 <211> 20  
 <212> DNA  
 <213> artificial sequence  
 <220>  
 <221> primer\_bind  
 <223> primer NFP used for isolation of fluorescent  
 protein  
 <400> 13

ttccayggtr tgaayttycc

20

<210> 14  
 <211> 4  
 <212> PRT  
 <213> *Aequorea victoria*  
 <220>  
 <222> 134...137  
 <223> amino acid sequence of a key stretch on which  
 <400> primers PVMa and PVMb are based  
 14

Gly Pro Val Met

<210> 15  
 <211> 21  
 <212> DNA  
 <213> artificial sequence  
 <220>  
 <221> primer\_bind



<222> 15

<223> primer PVM<sub>a</sub> used for isolation of fluorescent protein; n at position 15 represents any of the four bases

<400> 15

cctgccrayg gtccngtmat g

21

<210> 16

<211> 21

<212> DNA

<213> artificial sequence

<220>

<221> primer\_bind

<222> 15

<223> primer PVM<sub>b</sub> used for isolation of fluorescent protein; n at position 15 represents any of the four bases

<400> 16

cctgccrayg gtccngtakat g

21

<210> 17

<211> 47

<212> DNA

<213> artificial sequence

<220>

<221> primer\_bind

<223> primer T7-TN3 used in cDNA synthesis and RACE

<400> 17

gtaatacgac tcactatagg gccgcagtcg accgtttttt tttttttt

47

<210> 18

<211> 45

<212> DNA

<213> artificial sequence

<220>

<221> primer\_bind



<223> primer T7-TS used in cDNA synthesis and RACE  
<400> 18

gtaatacgac tcactatagg gcaaggcgtg gatatcaacgc agagt  
45

<210> 19  
<211> 22  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> primer T7 used in cDNA synthesis and RACE  
<400> 19

gtaatacgac tcactatagg gc 22

<210> 20  
<211> 21  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> gene-specific primer used for 5'-RACE for  
Anemonia  
majano  
<400> 20

gaaatagtca ggcatactgg t 21

<210> 21  
<211> 20  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> gene-specific primer used for 5'-RACE for  
Anemonia  
majano  
<400> 21



gtcaggcata ctggtaggat

20

<210> 22  
<211> 21  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> gene-specific primer used for 5'-RACE for  
*Clavularia* sp.  
<400> 22

cttgaaatag tctgctatat c

21

<210> 23  
<211> 19  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> gene-specific primer used for 5'-RACE for  
*Clavularia* sp.  
<400> 23

tctgctatat cgtctgggt

19

<210> 24  
<211> 23  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> gene-specific primer used for 5'-RACE for  
*Zoanthus*  
*sp.*  
<400> 24

gttcttgaaa tagtctacta tgt

23



<210> 25  
 <211> 20  
 <212> DNA  
 <213> artificial sequence  
 <220>  
 <221> primer\_bind  
 <223> gene-specific primer used for 5'-RACE for  
*Zoanthus* sp.  
 <400> 25

gtctactatg tcttgaggat

20

<210> 26  
 <211> 19  
 <212> DNA  
 <213> artificial sequence  
 <220>  
 <221> primer\_bind  
 <223> gene-specific primer used for 5'-RACE for  
*Discosoma* sp. "red"  
 <400> 26

caagcaaatg gcaaaggta

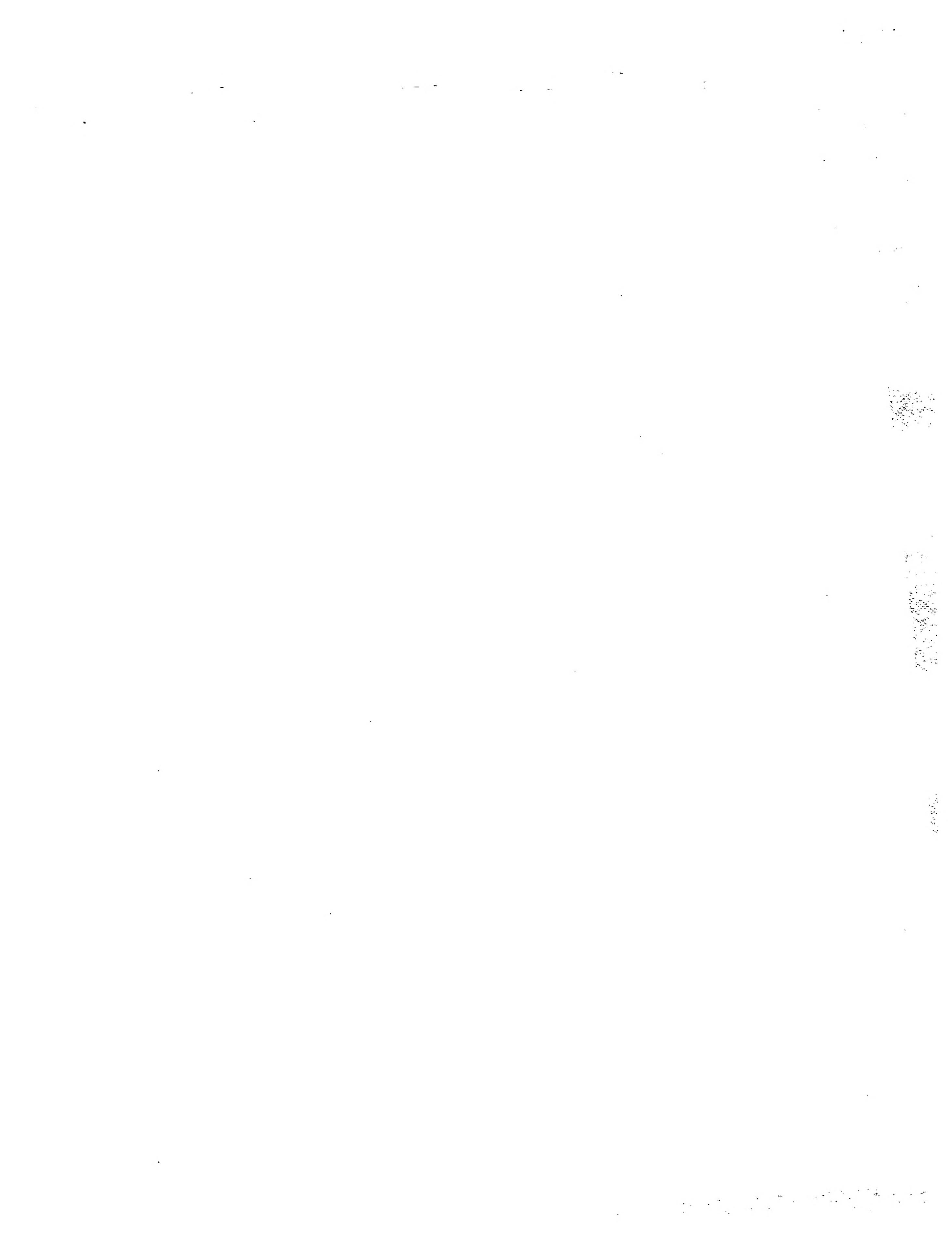
19

<210> 27  
 <211> 19  
 <212> DNA  
 <213> artificial sequence  
 <220>  
 <221> primer\_bind  
 <223> gene-specific primer used for 5'-RACE for  
*Discosoma* sp. "red"  
 <400> 27

cggattatgtg gccttcgtt

19

<210> 28



<211> 19  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> gene-specific primer used for 5'-RACE for  
*Discosoma striata*  
<400> 28

ttgtcttctt ctgcacaac 19

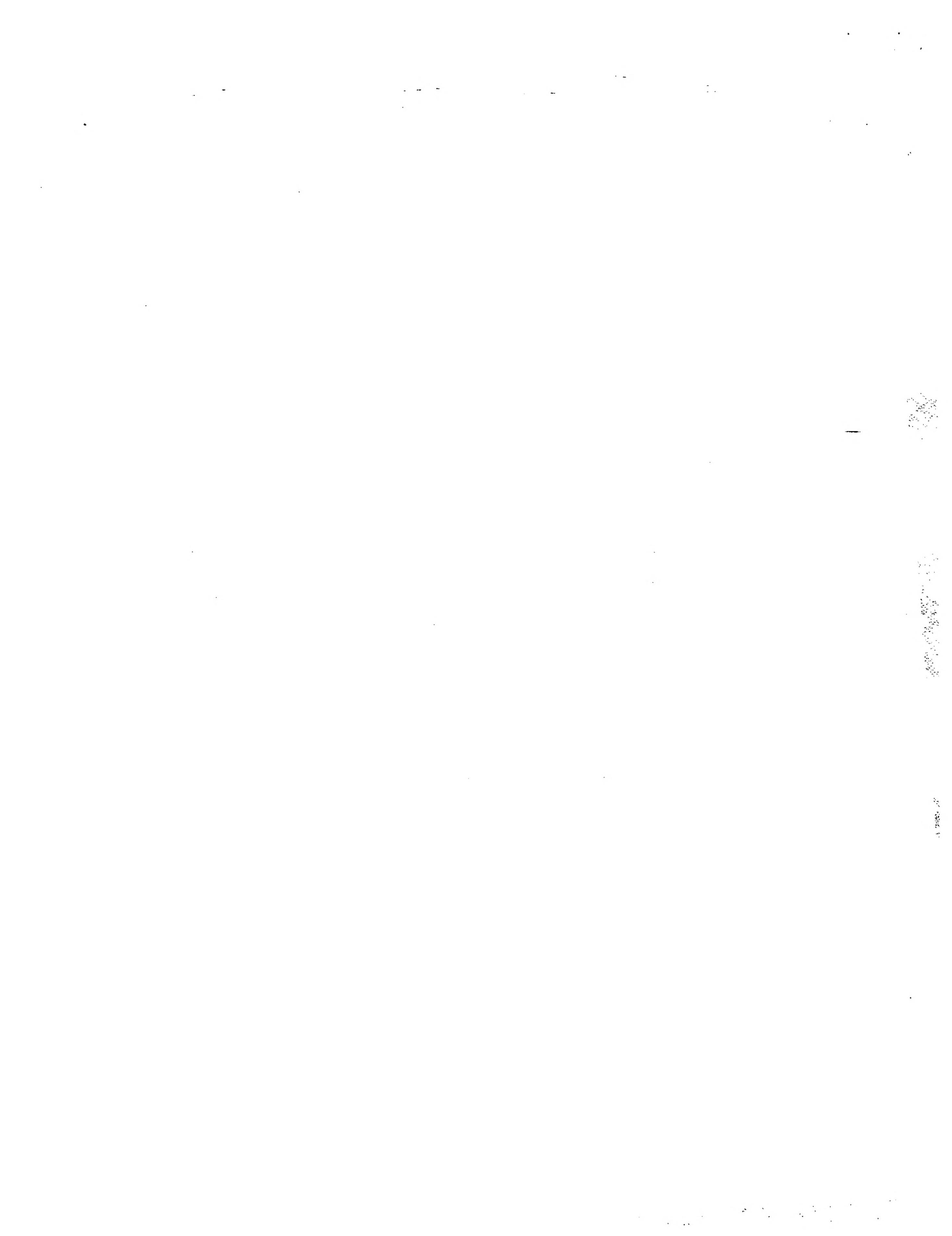
<210> 29  
<211> 17  
<212> DNA —  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> gene-specific primer used for 5'-RACE for  
*Discosoma striata*  
<400> 29

ctgcacaacg ggtccat 17

<210> 30  
<211> 20  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> gene-specific primer used for 5'-RACE for  
*Anemonia sulcata*  
<400> 30

cctctatctt catttcctgc 20

<210> 31  
<211> 20  
<212> DNA  
<213> artificial sequence



&lt;220&gt;

&lt;221&gt; primer\_bind

<223> gene-specific primer used for 5'-RACE for  
Anemonia

sulcata

&lt;400&gt; 31

tatcttcatt tcctgcgtac

20

&lt;210&gt; 32

&lt;211&gt; 19

&lt;212&gt; DNA

&lt;213&gt; artificial sequence

&lt;220&gt;

&lt;221&gt; primer\_bind

<223> gene-specific primer used for 5'-RACE for  
Discosoma sp. "magenta"

&lt;400&gt; 32

ttcagcaccc catcacgag

19

&lt;210&gt; 33

&lt;211&gt; 19

&lt;212&gt; DNA

&lt;213&gt; artificial sequence

&lt;220&gt;

&lt;221&gt; primer\_bind

<223> gene-specific primer used for 5'-RACE for  
Discosoma sp. "magenta"

&lt;400&gt; 33

acgctcagag ctgggttcc

19

&lt;210&gt; 34

&lt;211&gt; 22

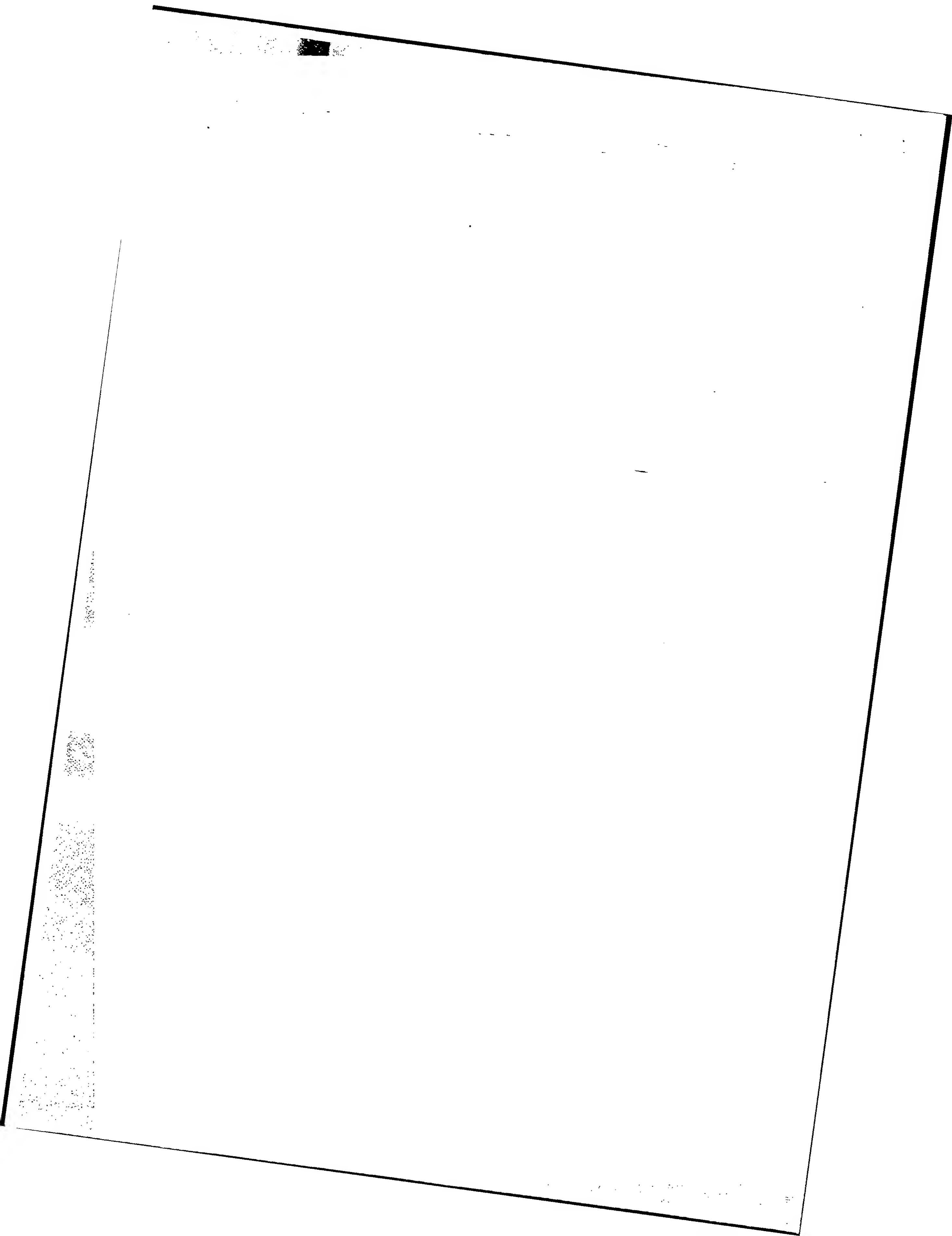
&lt;212&gt; DNA

&lt;213&gt; artificial sequence

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;223&gt; gene-specific primer used for 5'-RACE for



*Discosoma sp. "green"*

&lt;400&gt; 34

ccctcagcaa tccatcacgt tc 22

&lt;210&gt; 35

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; artificial sequence

&lt;220&gt;

&lt;221&gt; primer\_bind

<223> gene-specific primer used for 5'-RACE for  
*Discosoma sp. "green"*

&lt;400&gt; 35

attatctcag tggatggttc 20

&lt;210&gt; 36

&lt;211&gt; 31

&lt;212&gt; DNA

&lt;213&gt; artificial sequence

&lt;220&gt;

&lt;221&gt; primer\_bind

<223> upstream primer used to obtain full coding  
regionof nFPs from *Anemonia majano*

&lt;400&gt; 36

acatggatcc gctctttcaa acaagtttat c 31

&lt;210&gt; 37

&lt;211&gt; 34

&lt;212&gt; DNA

&lt;213&gt; artificial sequence

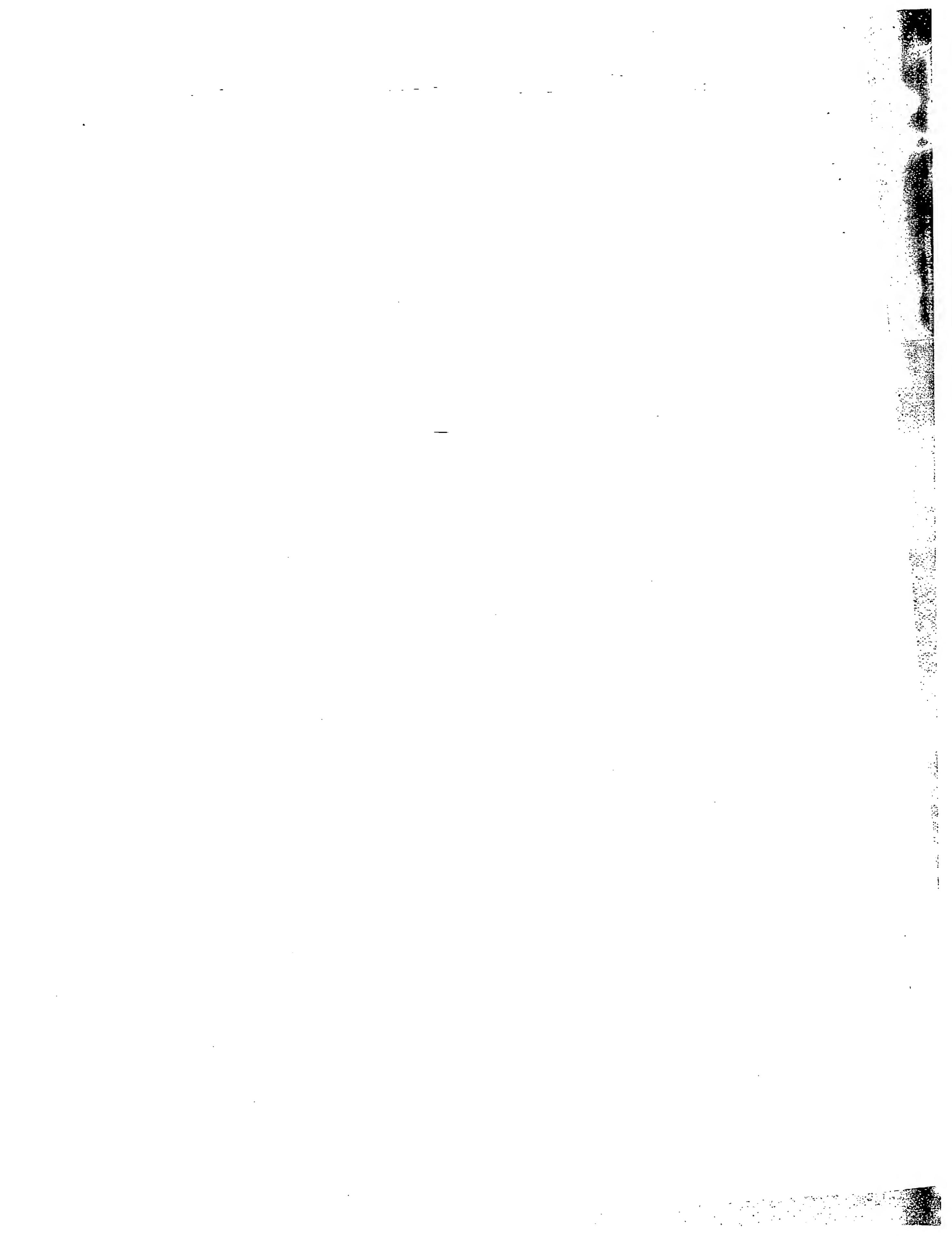
&lt;220&gt;

&lt;221&gt; primer\_bind

<223> downstream primer used to obtain full coding  
region of nFPs from *Anemonia majano*

&lt;400&gt; 37

tagtactcga gcttattcgt atttcagtga aatc 34



<210> 38  
<211> 29  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> upstream primer used to obtain full coding  
region of nFPs from *Clavularia* sp.  
<400> 38

acatggatcc aacattttt tgagaaaacg 29

<210> 39  
<211> 28  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> upstream primer used to obtain full coding  
region of nFPs from *Clavularia* sp.  
<400> 39

acatggatcc aaagctctaa ccaccatg 28

<210> 40  
<211> 31  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> downstream primer used to obtain full coding  
region of nFPs from *Clavularia* sp.  
<400> 40

tagtactcga gcaacacaaa ccctcagaca a 31

<210> 41



<211> 28  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> upstream primer used to obtain full coding  
region  
of nFPs from *Zoanthus sp.*  
<400> 41

acatggatcc gctcagtcaa agcacggt 28

<210> 42  
<211> 32  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> downstream primer used to obtain full coding  
region of nFPs from *Zoanthus sp.*  
<400> 42

tagtactcga ggttggaact acattcttat ca 32

<210> 43  
<211> 31  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> upstream primer used to obtain full coding  
region  
of nFPs from *Discosoma sp. "red"*  
<400> 43

acatggatcc aggtcttcca agaatgttat c 31

<210> 44  
<211> 29  
<212> DNA



<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> downstream primer used to obtain full coding  
region of nFPs from *Discosoma* sp. "red"  
<400> 44

tagtactcga ggagccaa gt tcagcctta

29

<210> 45  
<211> 28  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> upstream primer used to obtain full coding  
region of nFPs from *Discosoma striata*  
<400> 45

acatggatcc agttggtcca agagtgtg

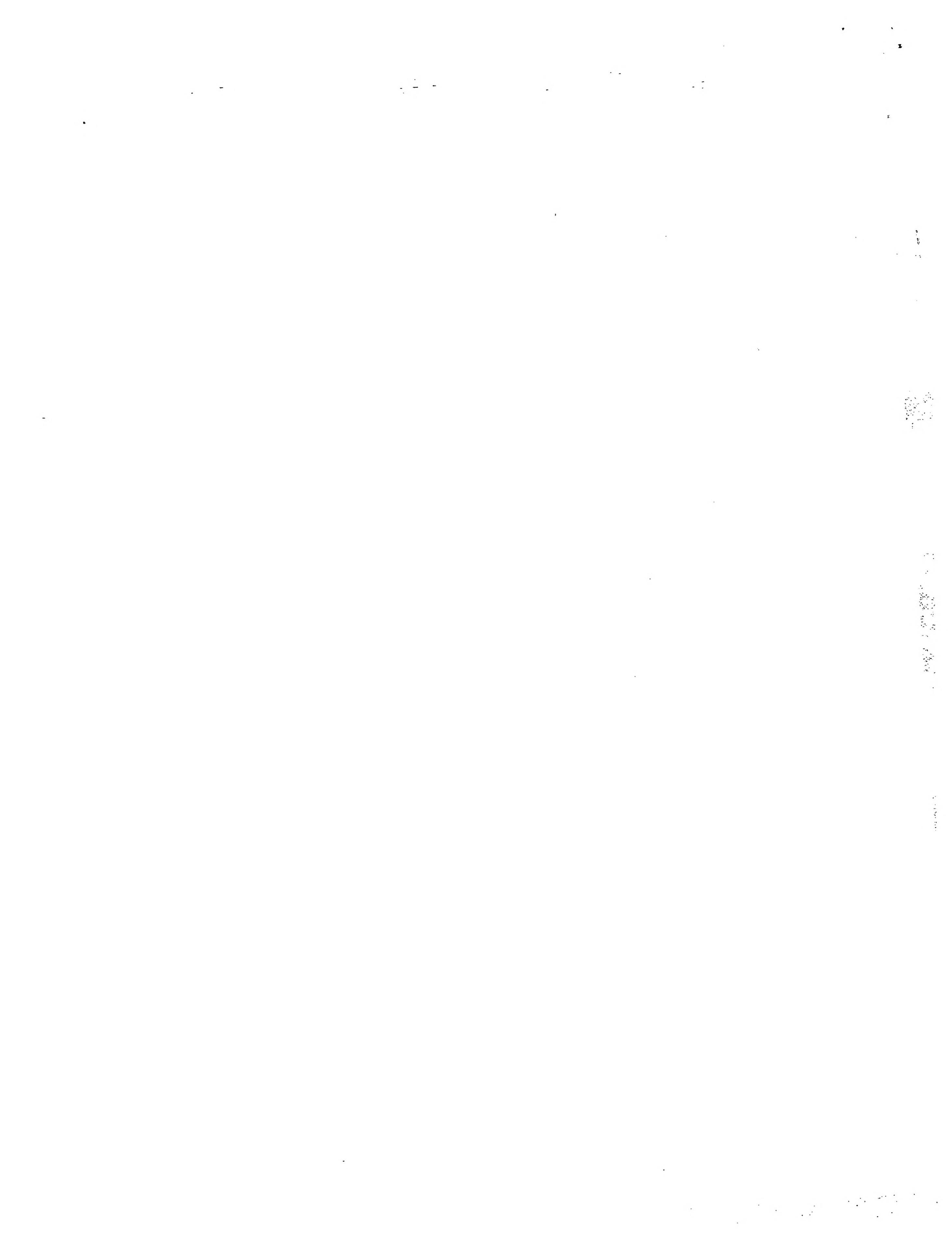
28

<210> 46  
<211> 28  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind  
<223> downstream primer used to obtain full coding  
region of nFPs from *Discosoma striata*  
<400> 46

tagcgagctc tatcatgcct cgtcacct

28

<210> 47  
<211> 31  
<212> DNA  
<213> artificial sequence  
<220>  
<221> primer\_bind



<223> upstream primer used to obtain full coding  
region of nFPs from *Anemonia sulcata*

<400> 47

acatggatcc gcttcctttt taaagaagac t 31

<210> 48

<211> 28

<212> DNA

<213> artificial sequence

<220>

<221> primer\_bind

<223> downstream primer used to obtain full coding  
region of nFPs from *Anemonia sulcata*

<400> 48

tagtactcga gtccttggga gcggcttg 28

<210> 49

<211> 30

<212> DNA

<213> artificial sequence

<220>

<221> primer\_bind

<223> upstream primer used to obtain full coding  
region of nFPs from *Discosoma sp. "magenta"*

<400> 49

acatggatcc agttgttcca agaatgtgat 30

<210> 50

<211> 26

<212> DNA

<213> artificial sequence

<220>

<221> primer\_bind

<223> downstream primer used to obtain full coding  
region of nFPs from *Discosoma sp. "magenta"*



<400> 50

tagtactcga ggccattacg ctaatc 26

<210> 51

<211> 31

<212> DNA

<213> artificial sequence

<220>

<221> primer\_bind

<223> upstream primer used to obtain full coding region of nFPs from *Discosoma* sp. "green"

<400> 51

acatggatcc agtgcactta aagaagaaat g 31

<210> 52

<211> 29

<212> DNA

<213> artificial sequence

<220>

<221> primer\_bind

<223> downstream primer used to obtain full coding region of nFPs from *Discosoma* sp. "green"

<400> 52

tagtactcga gattcggtt aatgccttg 29

<210> 53

<211> 33

<212> DNA

<213> artificial sequence

<220>

<221> primer\_bind

<223> TS-oligo used in cDNA synthesis and RACE

<400> 53

aagcagtggatcaacgcag agtacgcgrgr grg 33

<210> 54



<211>	238														
<212>	PRT														
<213>	<i>Aequorea victoria</i>														
<220>															
<223>	amino acid sequence of GFP														
<400>	54														
Met	Ser	Lys	Gly	Glu	Glu	Leu	Phe	Thr	Gly	Val	Val	Pro	Ile	Leu	
															15
5										10					
Val	Glu	Leu	Asp	Gly	Asp	Val	Asn	Gly	His	Lys	Phe	Ser	Val	Ser	
															30
20										25					
Gly	Glu	Gly	Glu	Gly	Asp	Ala	Thr	Tyr	Gly	Lys	Leu	Thr	Leu	Lys	
															45
35										40					
Phe	Ile	Cys	Thr	Thr	Gly	Lys	Leu	Pro	Val	Pro	Trp	Pro	Thr	Leu	
															60
50										55					
Val	Thr	Thr	Phe	Ser	Tyr	Gly	Val	Gln	Cys	Phe	Ser	Arg	Tyr	Pro	
															75
65										70					
Asp	His	Met	Lys	Gln	His	Asp	Phe	Phe	Lys	Ser	Ala	Met	Pro	Glu	
															90
80										85					
Gly	Tyr	Val	Gln	Glu	Arg	Thr	Ile	Phe	Phe	Lys	Asp	Asp	Gly	Asn	
															105
95										100					
Tyr	Lys	Thr	Arg	Ala	Glu	Val	Lys	Phe	Glu	Gly	Asp	Thr	Leu	Val	
															120
110										115					
Asn	Arg	Ile	Glu	Leu	Lys	Gly	Ile	Asp	Phe	Lys	Glu	Asp	Gly	Asn	
															135
125										130					
Ile	Leu	Gly	His	Lys	Leu	Glu	Tyr	Asn	Tyr	Asn	Ser	His	Asn	Val	
															150
140										145					
Tyr	Ile	Met	Ala	Asp	Lys	Gln	Lys	Asn	Gly	Ile	Lys	Val	Asn	Phe	
															165
155										160					
Lys	Ile	Arg	His	Asn	Ile	Glu	Asp	Gly	Ser	Val	Gln	Leu	Ala	Asp	
															180
170										175					
His	Tyr	Gln	Gln	Asn	Thr	Pro	Ile	Gly	Asp	Gly	Pro	Val	Leu	Leu	
															195
185										190					
Pro	Asp	Asn	His	Tyr	Leu	Ser	Thr	Gln	Ser	Ala	Leu	Ser	Lys	Asp	
															210
200										205					
Pro	Asn	Glu	Lys	Arg	Asp	His	Met	Val	Leu	Leu	Glu	Phe	Val	Thr	
															225
215										220					
Ala	Ala	Gly	Ile	Thr	His	Gly	Met	Asp	Glu	Leu	Tyr	Lys			
230										235					



<210> 55  
 <211> 919  
 <212> DNA  
 <213> *Discosoma sp. "green"*  
 <220>  
 <221> CDS  
 <223> cDNA sequence of dgFP512  
 <400> 55

attcacctcg gtgatttcta agagaaagga tcaccatcaa gagaagagct 50  
 gtaaaaagtta atattttact gtacttctac cagcatgagt gcacttaaag 100  
 aagaaaatgaa aatcaacctt acaatggaag gtgttgttaa cgggcttcca 150  
 tttaagatcc gtggggatgg aaaaggcaaa ccataccagg gatcacagga 200  
 gttaaccttg acggtggtta aaggcgggcc tctgccttcc tcttatgata 250  
 ttctgacaac gatgtttcag tacggcaaca gggcattcgt aaactaccca 300  
 gaggacatac cagatattt caagcagacc tgttctggtc ctaatggtgg 350  
 atattcctgg caaaggacca tgacttatga agacggaggc gtttgcactg 400  
 ctacaagcaa catcagcgtg gttggcgaca ctttcaatta tgacattcac 450  
 tttatggag cgaattttcc tcttgatggt ccagtgatgc agaaaagaac 500  
 aatgaaatgg gaaccatcca ctgagataat gtttgaacgt gatggaatgc 550  
 tgaggggtga cattgccatg tctctgttgc tgaagggagg gggccattac 600  
 cgatgtgatt ttgaaactat ttataaacc aataaggttg tcaagatgcc 650  
 agattaccat ttgtggacc actgcattga gataacgagt caacaggatt 700  
 attacaacgt ggtttagctg accgaggttg ctgaagcccg ctactcttcg 750  
 ctggagaaaa tcggcaaatac aaaggcgtaa atccaagcaa tctaagaaaa 800  
 caacaaggca ttaaaccgaa tcaccgttt gaattttcg ttggaaattt 850  
 cttggtaaaa ctaggttttag aacgtttcat ttgcgtggac ttcttgact 900  
 cagctgtaga caagaaaga 919

<210> 56  
 <211> 231  
 <212> PRT  
 <213> *Discosoma sp. "green"*  
 <220>  
 <223> amino acid sequence of dgFP512  
 <400> 56

Met Ser Ala Leu Lys Glu Glu Met Lys Ile Asn Leu Thr Met Glu  
 5 10 15



Gly Val Val Asn Gly Leu Pro Phe Lys Ile Arg Gly Asp Gly Lys  
20 25 30  
Gly Lys Pro Tyr Gln Gly Ser Gln Glu Leu Thr Leu Thr Val Val  
35 40 45  
Lys Gly Gly Pro Leu Pro Phe Ser Tyr Asp Ile Leu Thr Thr Met  
50 55 60  
Phe Gln Tyr Gly Asn Arg Ala Phe Val Asn Tyr Pro Glu Asp Ile  
65 70 75  
Pro Asp Ile Phe Lys Gln Thr Cys Ser Gly Pro Asn Gly Gly Tyr  
80 85 90  
Ser Trp Gln Arg Thr Met Thr Tyr Glu Asp Gly Gly Val Cys Thr  
95 100 105  
Ala Thr Ser Asn Ile Ser Val Val Gly Asp Thr Phe Asn Tyr Asp  
110 115 120  
Ile His Phe Met Gly Ala Asn Phe Pro Leu Asp Gly Pro Val Met  
125 130 135  
Gln Lys Arg Thr Met Lys Trp Glu Pro Ser Thr Glu Ile Met Phe  
140 145 150  
Glu Arg Asp Gly Met Leu Arg Gly Asp Ile Ala Met Ser Leu Leu  
155 160 165  
Leu Lys Gly Gly His Tyr Arg Cys Asp Phe Glu Thr Ile Tyr  
170 175 180  
Lys Pro Asn Lys Val Val Lys Met Pro Asp Tyr His Phe Val Asp  
185 190 195  
His Cys Ile Glu Ile Thr Ser Gln Gln Asp Tyr Tyr Asn Val Val  
200 205 210  
Glu Leu Thr Glu Val Ala Glu Ala Arg Tyr Ser Ser Leu Glu Lys  
215 220 225  
Ile Gly Lys Ser Lys Ala  
230



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US99/29404

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C07K 14/435; C12N 1/00, 1/19, 5/10, 15/12, 15/63

US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/320.1, 252.3, 252.33, 325, 410, 254.11, 348, 369, 69.1; 530/350; 536/23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
*****	The sequence diskette submitted with the description was defective; thus the documents listed below were obtained solely by a word search. No SEQ ID NOS. could be searched.	*****
X, P	MATZ et al. Fluorescent proteins from nonbioluminescent Anthozoa species. Nature Biotechnology. October 1999, Volume 17, No. 10, pages 969-673, entire document.	1-23
X, P	DE 197 18 640 A1 (WIEDENMANN) 22 July 1999, entire document.	14-15, 20

Further documents are listed in the continuation of Box C.  See patent family annex.

Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means		
"P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

07 MARCH 2000

Date of mailing of the international search report

06 APR 2000

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

GABRIELE ELISABETH BUGAISKY

Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US99/29404

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ANDERLUH et al. Cloning, sequencing, and expression of equinatoxin II. Biochemical and Biophysical Research Communications. 1996, Volume 220, No. 2, pages 437-442, entire document.	1-2, 9, 11-15, 20
X ---	MACEK et al. Intrinsic tryptophan fluorescence of equinatoxin II, a pore-forming polypeptide from the sea anemone, <i>Actinia equina</i> L, monitors its interaction with lipid membranes. European Journal of Biochemistry. 1995, Volume 234, pages 329-335, entire document. Cited as "L" document because it establishes fluorescence of equinatoxin II.	14-15, 20 -----
L		1-2, 9,11-13

**INTERNATIONAL SEARCH REPORT**

International application No. PCT/US99/29404
---

**A. CLASSIFICATION OF SUBJECT MATTER:**  
**US CL :**

435/320.1, 252.3, 252.33, 325, 410, 254.11, 348, 369, 69.1; 530/350; 536/23.5

**B. FIELDS SEARCHED**

Electronic data bases consulted (Name of data base and where practicable terms used):

Dialog files 155, 5, 434, 34, 357, 28, 44, 77 (Medline, biosis, Scisearch, Derwent Biotech. Abs., Oceanic Abs., Aquatic Sci. & Fish Abs., Dissertation Abs. Online, Conference Papers Index); STN-CAS files registry, CAPLUS; WEST files USPT, Derwent WPI

search terms: fluoresc?, bioluminesc?, protein?, polypeptide?, gene#, anthozo?, zoanhar?, corallimorph?, discosom?, cnidar?, inverteb?, anemone?, dgfp512, rhodact?, coral?, msalkeem/sqsp, vngh/sqep, gegeg/sqep, gegng/sqep, gmnfp/sqep, gvnpf/sqep, gpvm/sqep

